



IFW16

RAW SEQUENCE LISTING

DATE: 10/07/2004

PATENT APPLICATION: US/10/019,359

TIME: 13:45:04

Input Set : A:\PTO.FG.txt

Output Set: N:\CRF4\10072004\J019359.raw

3 <110> APPLICANT: K. U. Leuven Research & Development
 4 Debyser, Zeger
 5 De Clercq, Erik
 6 Cherepanov, Peter
 7 Pluymers, Wim
 9 <120> TITLE OF INVENTION: A synthetic gene for expression of a retroviral protein with
 wild type activity in eukaryotic cells
 10
 12 <130> FILE REFERENCE: K1291-PCT
 14 <140> CURRENT APPLICATION NUMBER: 10/019,359
 C--> 15 <141> CURRENT FILING DATE: 2002-04-11
 17 <150> PRIOR APPLICATION NUMBER: EP99201306.0
 18 <151> PRIOR FILING DATE: 1999-04-26
 20 <150> PRIOR APPLICATION NUMBER: EP00200171.1
 21 <151> PRIOR FILING DATE: 2000-01-18
 23 <160> NUMBER OF SEQ ID NOS: 2
 25 <170> SOFTWARE: PatentIn version 3.3
 27 <210> SEQ ID NO: 1
 28 <211> LENGTH: 930
 29 <212> TYPE: DNA
 30 <213> ORGANISM: Artificial sequence
 32 <220> FEATURE:
 33 <223> OTHER INFORMATION: description of artificial sequence: synthetic gene encoding
 HIV integrase
 34
 37 <220> FEATURE:
 38 <221> NAME/KEY: misc_signal
 39 <222> LOCATION: (24)..(30)
 40 <223> OTHER INFORMATION: Kozak sequence
 42 <220> FEATURE:
 43 <221> NAME/KEY: CDS
 44 <222> LOCATION: (27)..(899)
 46 <400> SEQUENCE: 1
 47 atcactagca acctcaaaca gacacc atg gga ttc ctg gac ggc att gac aag 53
 48 Met Gly Phe Leu Asp Gly Ile Asp Lys
 49 1 5
 51 gct cag gag gag cac gag aag tac cac tcg aat tgg cgg gcc atg gcc 101
 52 Ala Gln Glu Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala
 53 10 15 20 25
 55 tcc gac ttc aac ctg cca ccc gtc gtc gct aag gag atc gtt gct agc 149
 56 Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser
 57 30 35 40
 59 tgc gac aag tgc cag ctg aaa ggc gag gct atg cac ggg cag gtt gat 197
 60 Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp

61

45

50

55

RAW SEQUENCE LISTING

DATE: 10/07/2004

PATENT APPLICATION: US/10/019,359

TIME: 13:45:04

Input Set : A:\PTO.FG.txt

Output Set: N:\CRF4\10072004\J019359.raw

```

63 tgc tct ccc ggc atc tgg cag ctc gac tgt act cac ctg gag ggc aag      245
64 Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys
65          60          65          70
67 gtc atc ctg gtc gcc gtg cac gtg gcc tct ggt tac atc gag gct gag      293
68 Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu
69      75          80          85
71 gtc atc cct gca gag act ggc cag gag act gcc tat ttc ctg ctg aaa      341
72 Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys
73 90          95          100          105
75 ctg gcc ggc cgg tgg cct gtg aag aca gtg cac aca gat aac ggc tcc      389
76 Leu Ala Gly Arg Trp Pro Val Lys Thr Val His Thr Asp Asn Gly Ser
77          110          115          120
79 aac ttc acc tcc acc act gtg aag gct gcc tgc tgg tgg gct ggg atc      437
80 Asn Phe Thr Ser Thr Thr Val Lys Ala Ala Cys Trp Trp Ala Gly Ile
81          125          130          135
83 aag cag gag ttc ggg atc ccc tat aac cca cag tct cag ggc gtg atc      485
84 Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Ile
85          140          145          150
87 gaa tcc atg aac aag gag ctg aag aag atc atc ggc cag gtt cgg gac      533
88 Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp
89      155          160          165
91 cag gca gag cac ctg aag act gca gtg cag atg gcc gtg ttc atc cac      581
92 Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His
93 170          175          180          185
95 aac ttc aag cga aag ggc ggc atc ggt ggc tac tca gcc ggc gag cgg      629
96 Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg
97          190          195          200
99 atc gtg gac atc atc gcc act gac atc cag acc aaa gag ctg cag aag      677
100 Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys
101          205          210          215
103 cag atc acc aag atc cag aac ttc cgt gtg tac tac cgg gac tcc cgg      725
104 Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg
105          220          225          230
107 gac cct gtg tgg aag ggc cct gcc aag ctg ctg tgg aag ggc gag ggc      773
108 Asp Pro Val Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly
109      235          240          245
111 gcc gtg gtc att cag gac aac tct gac atc aag gtt gtg ccc agg cgc      821
112 Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg
113 250          255          260          265
115 aag gcc aag att atc cgg gac tac ggc aag cag atg gct ggc gac gac      869
116 Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp
117          270          275          280
119 tgt gtg gcc tct cgt caa gat gag gac taa gtc caactac taaactgggg      919
120 Cys Val Ala Ser Arg Gln Asp Glu Asp
121          285          290
123 gatattatga t      930
127 <210> SEQ ID NO: 2
128 <211> LENGTH: 290
129 <212> TYPE: PRT

```

RAW SEQUENCE LISTING

DATE: 10/07/2004

PATENT APPLICATION: US/10/019,359

TIME: 13:45:04

Input Set : A:\PTO.FG.txt

Output Set: N:\CRF4\10072004\J019359.raw

130 <213> ORGANISM: Artificial sequence

132 <220> FEATURE:

133 <223> OTHER INFORMATION: Synthetic Construct

135 <400> SEQUENCE: 2

```

137 Met Gly Phe Leu Asp Gly Ile Asp Lys Ala Gln Glu Glu His Glu Lys
138 1          5          10          15
141 Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
142          20          25          30
145 Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
146          35          40          45
149 Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
150          50          55          60
153 Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
154 65          70          75          80
157 Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
158          85          90          95
161 Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
162          100          105          110
165 Lys Thr Val His Thr Asp Asn Gly Ser Asn Phe Thr Ser Thr Thr Val
166          115          120          125
169 Lys Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
170          130          135          140
173 Tyr Asn Pro Gln Ser Gln Gly Val Ile Glu Ser Met Asn Lys Glu Leu
174 145          150          155          160
177 Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
178          165          170          175
181 Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly
182          180          185          190
185 Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
186          195          200          205
189 Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
190          210          215          220
193 Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro Val Trp Lys Gly Pro
194 225          230          235          240
197 Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
198          245          250          255
201 Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
202          260          265          270
205 Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
206          275          280          285
209 Glu Asp
210          290

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/019,359

DATE: 10/07/2004

TIME: 13:45:05

Input Set : A:\PTO.FG.txt

Output Set: N:\CRF4\10072004\J019359.raw

L:15 M:271 C: Current Filing Date differs, Replaced Current Filing Date